COMMUNICATION AND STRATEGIC INFERENCE*

1. Introduction

In this paper I give an account of Gricean communication using the frameworks of situation theory and game theory. I develop informally both a model of communication and an approximate set of necessary and sufficient conditions for communication. I will first describe the problem in broad terms and then provide a detailed analysis of a single example of communication. I will not discuss here the more basic issues connected with Grice's attempt to distinguish natural from nonnatural meaning. Instead, my strategy will be to focus first on how communication is possible once a language is given. Many aspects of this problem carry over to the general problem of communication. My attempt is to develop tools that will provide an approach to this problem.

2. Meaning and Content

A language is a complex social institution. It is arguable that its primary function is communication. Indeed, it is possible to see language as arising from the communicative needs of a group of interacting agents. Lewis [28], for example, has argued for such a view. This makes communication a key concept in any account of language. The framework of situation theory and situation semantics developed by Barwise and Perry [14] and Grice's [19], [21] ideas on nonnatural meaning provide the best starting point for constructing a model of communication.

A trifle paradoxically perhaps, the central fact about most languages is that they are situated. This makes it possible for different propositions to be communicated in different circumstances with the same sentence. This is part of what makes language an efficient system of communication. Situation theory makes this context-dependence an integral part of utter-

---

* I am deeply indebted to my advisors Jon Barwise, Kenneth Arrow, and Ennio Stacchetti for their help with my dissertation and this paper which is based on a chapter of the dissertation. I would also like to thank R. Narasimhan, Rohit Parikh and Stephen Schiffer for many stimulating conversations. My dissertation was done in the Engineering-Economic Systems Department and at CSLI (Stanford University), and the paper at the Tata Institute of Fundamental Research, and I am indebted to all three institutions for their support.

ances. Once we allow situations a role in the determination of content, it becomes clear that there are certain aspects of utterances that are constant across utterances and others that vary from one situation to another. The most salient linguistic constant is what Barwise and Perry call the meaning of a sentence. This is difference from its content in an utterance. We could in fact write a simple schematic equation of the form \((\text{meaning of}) \text{ sentence} + \text{discourse situation} = \text{content}\).

Part of the task of a theory of communication is to explain how the equation above comes about. If we assume that a language is given that the problem is to show how we can get from meaning to content. Solving this problem in a completely general way turns out to be an extremely difficult task.

It is possible to take a step in this direction by bringing in the ideas of game theory and equilibrium developed by von Neumann, Nash, Arrow, Debreu, Aumann and other game theorists and economists. The ideas of rational agency and strategic interaction developed in this tradition provide the remaining part of the framework we need to solve this problem. They allow us to refine the schematic equation above to \(\text{agent architecture} + \text{sentence} + \text{situation of utterance} = \text{content}\). Grice’s important insights into these issues and what suggest that game-theoretic methods are the right ones to consider in looking at communication. This is further reinforced by Austin’s work, in particular, by his emphasis on seeing language as action (see [6], [7]).

Perhaps the key idea that we get from game theory is the concept of a strategic inference. Grice, and subsequently, Strawson [40], Lewis [28], and Schiffer [35], have shown how communication involves an extremely complex interplay of inferences about speaker’s and addressee’s intentions. It turns out that game theory, viewed broadly, provides a mathematical framework for precisely these kinds of inferences. However, game theory as currently formulated does not provide a readymade tool to model communication. It is necessary to develop its insights from first principles, more or less, keeping in mind its use in information economics.

3. Communication

Assume as given a large situation of environment \(\mathcal{E}\) with two rational agents \(A\) and \(B\) who share a language \(L\). \(A\) utters an indicative sentence \(\varphi\) assertively in an utterance situation \(u\) in order to communicate some information \(p^*\) to \(B\). On receiving \(A\)’s signal, \(B\) attempts to interpret \(A\)’s utterance in interpretive situation \(i\). \(A\) and \(B\) assume common knowl-
edge of their rationality and assume, moreover, that their interaction is a cooperative one.

Some aspects of the utterance will be public (see Barwise [8]). Typically, the sentence uttered will be publicly available to both agents after the utterance. Other aspects will, in general, be private, like the beliefs and intentions of the speaker and addressee. And there may be some aspects of the utterance that are partially shared by $A$ and $B$.

The general problem is to find the necessary and sufficient conditions (which involves finding the inferential mechanism) for $A$ to communicate some proposition $p^*$ to $B$. What sentence should $A$ choose to utter and how should $B$ interpret this choice of utterance, given their respective goals and information? The shared language $L$ provides shared meanings. Thus, the problem is to show how $A$ and $B$ can move jointly from meaning to content.

4. Strategic Inference

We can embed this problem in the larger picture of flows of information developed by Dretske [18] and Barwise and Perry [14]. As Barwise and Perry have argued, reality can be viewed as consisting of situations linked by a network of constraints. It is the constraint between two situations that makes one situation carry information about (naturally mean, in Grice's sense) another situation. A smoky situation involves a situation with something on fire. This is the constraint we describe when we say 'Smoke means fire'. An agent who sees the first situation and who knows the constraint can infer the existence of the second situation. And this agent can pass on this newly acquired information to another agent. Given a group of agents, or distributed system, there will be all kinds of transfers of information. A communication is a special type of information flow between agents.

The Gricean approach to communication suggests that a transfer of information between two agents will be communicative when both the proposition $p^*$ and the speaker's intention to convey it become public as a result of the interaction. This involves both the speaker and addressee jointly inferring various things about each other. I will call this joint two-sided inference a strategic inference.

My central claim is that all transfers of information between agents involve a strategic interaction between them. When the strategic interaction is common knowledge between the agents, that is, when it is a game (with a unique solution), the flow will be communicative.

I argue this by developing a detailed account of one strategic inference.
in isolation. Any complete utterance involves many separate acts and strategic inferences. For example, part of a communication will typically involve a referential act and the communication of this reference to the addressee. Each bit of information communicated will require its own strategic inference(s). Thus, any complete utterance involves an entire system of simultaneous strategic inferences. These inferences have to be simultaneous in general because they codetermine each other in general. For example, an utterance of ‘Mary had a little lamb’ will require inferring the designata of each of the five words in the sentence, (not to mention its internal structure), in order to determine the proposition expressed. No word has any particular priority in this determination. That is, there may be interactions among the various strategic inferences. And the embedding circumstances play a vital role in each inference. Mathematically, this amounts to a system of simultaneous equations.

I will focus on just one strategic inference by assuming that the addressee has available the partial information obtained from all the other inferences involved in interpreting an utterance. Thus, the addressee’s problem will be to use this partial information together with the utterance itself to get to the intended content.

As an example, I will consider a familiar kind of ambiguity, typically viewed as an ambiguity between two possible quantifier orderings, and see under what circumstances it might be disambiguated. A successful strategic inference requires a number of assumptions about the rationality of the agents involved, their goals and intentions, and their knowledge and beliefs. Not only that, it also involves important assumptions about the language they use, in particular, the kinds of choices the language affords. An important consequence of the analysis is that the content communicated will depend not only on what was uttered but also, crucially, on what the speaker might have uttered but chose not to and on their shared information about these choices.

I will build up the structure of a strategic inference step by step. This will make the role of the assumptions clearer and suggest ways in which the construction can be generalized or modified to include other complexities omitted here. The constructed Strategic Discourse Model (or SDM) turns out to be a new kind of game that I call a game of partial information. The content communicated will then be given by the Pareto-undominated Nash equilibrium of the game.

Any communication involves at least two things. One is its propositional content and the other is the illocutionary force with which this content is conveyed. I will first deal exclusively with the content and then extend the model to include the force of the utterance. This involves a consider-
ation of the appropriate sort of response by the addressee. And with this we get a slightly more complex game that is in certain respects similar to the signalling game studied in information economics.

I conclude with a partial account of the necessary and sufficient conditions for communication. This requires a consideration of more general strategic interactions that have been studied in game theory.

An essential part of the analysis is that this game-theoretic structure is situated somewhere, either in the 'minds' of agents or in their ambient circumstances or both. This makes the flow of information a situated communication. An important consequence of this situatedness is that agents are not required to have the potentially infinite set of nested intentions proposed by Grice and others. As Perry [33] has argued in the case of belief, the circumstantial nature of action does not require an agent to have every relevant belief explicitly present in its mind. We do not need to consider gravity each time we reach for a glass. In exactly the same way, communicating agents do not need to consider all the relevant intentions explicitly. The ambient game does much of the work. The view of communication, as a situated game, resolves a fair bit of the mystery associated with Gricean analyses of nonnatural meaning regarding how agents with quite finite capacities could ever hope to communicate if it required such inaccessibly complex intentions.

This covers most of the key points of my approach to the general problem of communication, and to the problem of how, given a shared language, communicating agents can get from meaning to content.

5. Communication and Strategic Inference

Suppose $A$, after having picked up the information in a recent report issued by The Mugges Association of New York (i.e. M.A.N.Y.), says to $B$ in situation $d$:

"Every ten minutes a man gets mugged in New York." ($\varphi$)

What does $A$ communicate to $B$ by this utterance (if anything) and how does this communication take place? $A$ could mean either that a particular man gets mugged every ten minutes (call this proposition $p'$) or that some person or other gets mugged every ten minutes (call this $p$). There are, of course, other ways in which $\varphi$ is ambiguous (e.g. 'every ten minutes', 'man', 'in', 'New York'), but our interest is in the ambiguity between $p$ and $p'$. So I assume that it is common knowledge between them that $p \lor p'$ is available to $B$ once $\varphi$ is uttered. I will call this disjunctive proposition the minimal content of $\varphi$ in $d$. To avoid considering
implicatures I will also assume that $A$ intends to give information about mugging to $B$ and that this is common knowledge between them.

Intuitively, given the circumstances above, we would be inclined to go with the second interpretation. It is difficult to imagine a man as immune to experience as would be required for the truth of $p'$. But this much merely tells us that (in the absence of other relevant shared information to the contrary) the second reading is the one more likely to be true. Under what conditions can $B$ select $p$ over $p'$ as the intended content with complete certainty?

First, note that there are circumstances in which $p'$ rather than $p$ might be the natural content of an utterance of $\varphi$. For example, $A$ might follow up with ‘He was interviewed on TV last night.’ If part of such a discourse, $B$ would have to interpret $A$'s first utterance as conveying $p'$.

With respect to $d$, however, it seems plausible to say that $B$ would infer $p$ as the intended content with certainty. In fact, we could say that $A$ communicates $p$ to $B$. That is, not only is $B$ able to infer $p$ from the utterance, but the structure of their interaction allows him to infer $A$'s intention to communicate $p$. It also allows $A$ to infer that $B$ can infer $p$ and that $B$ can infer $A$'s intention to communicate $p$. In fact, the structure of their interaction allows both $A$ and $B$ to infer full public information of $p$ and $A$'s intention to communicate $p$, all with complete certainty. That this suffices for communication is not obvious, and I will show later why it does.

I will make five sets of assumptions to explain this disambiguation and its communication. The first set applies to all situations of interest, more or less. The second applies to all the situations considered but they are not quite so general as those in the first set. The third involves more specific circumstantial assumptions. The fourth set contains certain provisional assumptions that would be unnecessary in a more complete model. The last set I mention only to point out that some such assumptions are needed, without worrying about the details.

For the first set, called the Communication Assumptions, assume that both $A$ and $B$ are rational agents. This means in essence that they choose actions that realise their goals in the best possible way, or in choice-theoretic terminology, they maximize their expected payoffs.

Next, in the absence of contrary information, agents can be assumed to prefer more relevant information to less, and to minimize the effort involved in realizing a goal, other things being equal. It may seem that this is derivable from rationality. In fact, it cannot be: the axioms of rationality admit an extremely broad range of preferences as rational. The first part of this assumption can however be derived from a more general
assumption, (though not one quite so general as rationality), and I will consider this later. This assumption restricts the range of payoffs agents try to maximize.

The third assumption is that agents cooperate in their communicative efforts. This is a further restriction on admissible payoffs. It entails that there is some higher-level goal the agents share and that they try to align their preferences with this goal, given what they know.

Finally, assume that $\mathcal{A}$ and $\mathcal{B}$ share a sufficiently expressive language $\mathcal{L}$, one that allows many ordinary propositions to be expressed “minimally”. (This is similar to Austin’s [7] notion of the explicit performative and to Searle’s [37] assumption of the existence of a literal paraphrase for a metaphor.)

All four assumptions are common knowledge.

There are four “intermediate” assumptions called the Information Assumptions. The first assumption is that the speaker is constrained to consider only sentences whose minimal content is true and relevant, (or at least believed to be true and relevant). It is certainly possible in general for speakers to say (more precisely, to express “minimally”) false or irrelevant things (knowingly). Many interesting implicatures arise in this way. (Irony: “He is a friend”. ) This assumption helps to restrict the domain of inquiry for now.

Next, the intended content is also assumed to be true and relevant. Correspondingly, the addressee is restricted to consider valid and relevant inferences given what he knows or believes.

All three assumptions are common knowledge (or at least mutual belief) between them.

The Circumstantial Assumptions contain, in this particular example, the assumption that $\mathcal{A}$ is attempting to convey some information to $\mathcal{B}$ about mugging in NYC and that this is common knowledge. This rules out the possibility of implicated contents. $\mathcal{A}$ also has the more specific intention to communicate $p$ to $\mathcal{B}$ and this is not known to $\mathcal{B}$. Also, $p'$ is unlikely, and expressing $p$ unambiguously takes greater effort than expressing it ambiguously and this is common knowledge.

The fourth set, the Minimal Content Assumptions, contain the assumption that it becomes common knowledge after the utterance that the minimal content of $\varphi$ in $d$ is $p \lor p'$. This enables us to focus on a single strategic inference.

The next provisional assumption is that it also becomes common knowledge that $\mathcal{A}$ intends to convey either $p$ or $p'$ exclusively. Note that this does not follow from the first assumption that the minimal content is publicly available. $\mathcal{A}$ could, for example, be wanting to convey the disjunc-
tion \( p \lor p' \) itself (deliberate ambiguity) or perhaps the conjunction \( p \land p' \) (as with puns, for example).

The four sets of assumptions taken together will be called the CICM assumptions. In any real communication there will be other features of the utterance and its interpretation that are relevant to the flow. \( A \) and \( B \) have to be located somewhere, their perceptual situations need to be specified, and their mode of communication has to be considered. Note that \( A \) and \( B \) need not be persons. For our purposes, it is enough to assume that all such required conditions are satisfied, whatever they are.

My claim then is that if all the five sets of conditions above are satisfied \( A \) will communicate \( p \) to \( B \).

6. GAMES RATIONAL AGENTS PLAY

There are at least four difficulties in analyzing a strategic inference. First, strategic inferences just are very complex inferences requiring many premises for their validity. Second, the process is a nonwellfounded one (see Aczel [1], Barwise [12]). Each agent has to consider what the other agent might do in order to figure out a rational course of action. This requires each agent to consider the other agent's taking into account the first agent's possible actions, and so on, ad infinitum. I do not consider this nonwellfoundedness explicitly here, but it will be clear that it is essential.

Thirdly, the model I develop is different from the traditional games of incomplete information. So though I use the tools of game theory the model is not a straightforward application of these tools and requires us to build a new type of game, a game of partial information, from scratch.

And then there is the problem of how to interpret the game-theoretic model. It would be plainly absurd to claim that people explicitly carry out the "steps" involved in a strategic inference. It seems better to see the game of partial information as a model of a class of constraints that captures the underlying logic of language and communication. Just as modus ponens describes the conditions for a legitimate deductive inference in traditional, "unsituated" logic without implying that an agent necessarily uses it explicitly to arrive at a warranted conclusion, so the SDM describes the structure of a valid strategic inference in a situated logic (Barwise, [10]) without implying anything about the actual procedures used by agents to arrive at the correct interpretation of an utterance. However, I will often talk as if agents are actually performing the various tasks involved because this makes the model more accessible from an intuitive point of view. Besides, not all aspects of the steps are dispensable, as we will see.
To start with, there is a discourse situation $d$ in which the information flow takes place. We can extract two component situations $u$ and $i$ from $d$. The utterance situation $u$ is the situation in which $A$ utters $\varphi$, and the interpretive situation $i$ is where $B$ attempts to interpret $A$'s utterance. In a more complete model we would need a third situation $r$, the reception situation, in which the addressee perceives the utterance.

If the attempted communication is successful, $B$ will interpret $\varphi$ in $d$ as $p$. I will assume that a situation $B$ called the background is embedded in every situation of interest (and $d$ in particular), and that $d$ is part of a larger situation $E$ called the environment, in which all events of interest occur. The Communication and Information Assumptions are assumed to hold in $B$ and the Circumstantial Assumptions in $d$.

I will first consider why $A$ chooses to say $\varphi$. This is one side of the coin whose other side turns out to be an account of why $B$ chooses to interpret $\varphi$ in $u$ as $p$. It is this interdependent choice that will be seen to constitute the joint act of communication.

The natural place to start in order to consider $A$'s choice of $\varphi$ is with $A$'s goal. She wants to communicate $p$ to $B$. Presumably, $L$ provides many sentences that will do the job in $d$. One of them may be $\varphi$. The minimal content of $\varphi$ is certainly both true and relevant to $A$'s goal. (M.A.N.Y. never exaggerates.) Since $L$ is expressive $A$ could also choose a sentence $\mu$ whose minimal content in $d$ is $p$. This might be a sentence like “Every ten minutes some man or other gets mugged in New York”. $A$ has thus at least two ways to attempt to communicate $p$. We could imagine others and collect them all together in what is called $A$’s choice set. This set of actions is known, strictly speaking, only to $A$. Of course, $B$ would know that it must be some subset of $L$, but this is obviously not very helpful (especially since $L$ is expressive.) In general, $B$ cannot know what $A$ believes to be true and relevant. Besides, we do not need to assume that $A$ considers all sentences that minimally express propositions believed true and relevant. $A$ is a finite agent and considers only small subsets of its possible choice set. So let's say that $A$ has a choice set that contains $\varphi$, $\mu$ and perhaps other sentences that could be used to express true and relevant propositions (minimally and otherwise).

As a rational agent $A$ needs to evaluate the possible consequences of each choice and then choose the best course of action. I will first consider the possible consequences of uttering $\varphi$. $B$ has to choose a proposition as the intended interpretation of $\varphi$ in $u$. The assumptions imply that $A$'s utterance raises the issue for $B$ whether $p$ or $p'$ is being communicated. $B$ has no prior way of settling this issue, given what he knows. This situation can be modelled by the graph in Figure 1.
The two initial nodes $s$ and $s'$ represent the two possible ways of settling the issue of whether $A$'s intention is to communicate $p$ or $p'$. Each initial node can be thought of as a situation that contains the relevant state of affairs together with other facts. Thus, in $s$ we have $A$ intending to communicate $p$, and in $s'$ we have her intending to communicate $p'$. Note that $A$ knows she is in $s$ and not in $s'$. But all $B$ knows, and knows only after $A$ has uttered $\varphi$, is that $A$ could be either in $s$ or in $s'$. The branches at the first level of the tree represent this action by $A$. Performing this action in $u$ gets $A$ from $s$ to $t$ and performing it in a corresponding hypothetical situation $u'$ gets $A$ from $s'$ to $t'$. Once again, $A$ knows that uttering $\varphi$ would result in $t$, but $B$ is unable to distinguish between $t$ and $t'$. The inability on $B$'s part is represented by an oval and $\{t, t'\}$ is called an "information set" for $B$. This is $B$'s situation after he perceives $\varphi$.

The Circumstantial Assumption that it is common knowledge that $p$ is more likely than $p'$ can, in the absence of any further countervailing information, justify the agents' taking it to be common knowledge between them that $A$ is probably attempting to communicate $p$ rather than $p'$. This shared knowledge that $s$ is more likely than $s'$ can be represented by the probabilities 0.9 and 0.1 respectively. Note that, in general, there is a big difference between the likelihood of a proposition's being true and the likelihood of an agent intending to communicate that proposition. It is the absence of further relevant information that justifies identifying the two probabilities.\(^1\)

$B$ has to choose an appropriate interpretation at this point. He has the same two choices at $t$ and $t'$ as shown in Figure 2, either $p$ or $p'$.

---

\(^1\) I have not specified the universe of propositions in which $p'$ is unlikely. It cannot be $\{p, p'\}$ because this set comes to be common knowledge only after $A$ has uttered $\varphi$. $A$'s utterance gives $B$ the information that either $p$ or $p'$ is being communicated (by the third Minimal Content Assumption above). Then $B$ can infer that, of the two, $p'$ is much less likely to be true than $p$. That is, $B$ can derive conditional probabilities for $p$ and $p'$, given the proposition $p \lor p'$. The underlying universe over which this conditioning is done will be some larger (unspecified/indefinite) set of propositions. Once $B$ gets conditional probabilities for $p$ and $p'$, he can, in the absence of further relevant information, more or less identify these probabilities which the corresponding conditional probabilities that $A$ is attempting to com-
Communication and Strategic Inference

Table 1. Summary of Assumptions

Communication Assumptions
1. \( \mathcal{A}, \mathcal{B} \) are rational.
2. \( \mathcal{A}, \mathcal{B} \) prefer more relevant information to less, and less effort to more, ceteris paribus.
3. \( \mathcal{A}, \mathcal{B} \) cooperate.
4. \( \mathcal{A}, \mathcal{B} \) share an expressive language \( \mathcal{L} \).
5. All of the above are common knowledge.

Information Assumptions
1. \( \mathcal{A} \) considers only those utterances whose minimal contents are true and relevant.
2. The intended content of \( \mathcal{A}'s \) utterance is true and relevant.
3. \( \mathcal{B} \) considers only valid and relevant inferences.
4. All of the above are common knowledge.

Circumstantial Assumptions
1. \( \mathcal{A} \) is attempting to communicate \( p \).
2. \( \mathcal{A} \) is attempting to communicate some information about mugging to \( \mathcal{B} \).
3. \( p' \) is unlikely.
4. Expressing \( p \) unambiguously takes greater effort than expressing it ambiguously.
5. \( \mathcal{A} \) utters \( \varphi \) in \( d \).
6. \( \mathcal{B} \) does not know (1). The rest are common knowledge.

Minimal Content Assumptions
1. The minimal content of \( \varphi \) in \( d \) is \( p \lor p' \).
2. \( \mathcal{A} \) is attempting to communicate either \( p \) or \( p' \).
3. The above become common knowledge after the utterance.

Unspecified Assumptions
These include all required assumptions about perception, processing, and the mode of communication.

Since \( \mathcal{B} \) knows that either \( t \) or \( t' \) is factual, he can infer that \( \mathcal{A} \) intends to communicate either \( p \) or \( p' \) (or equivalently, that \( s \) or \( s' \) is factual). If he is in \( t \), \( p \) is the preferred choice, and if he is in \( t' \), \( p' \) is the preferred choice. Unlike standard game theory however, these preferences are not externally given but derived internally from the relevant intentions, exactly communicate \( p \) or to communicate \( p' \), given that \( \mathcal{A} \) is attempting to communicate either \( p \) or \( p' \). Finally, all this is common knowledge, getting us to the represented probabilities above.
as required by Grice. That is, if he is in \( t \), it is possible to infer that \( s \) is factual and that \( \mathcal{A} \) intends to communicate \( p \). So, in \( t \), \( \mathcal{B} \) can recognize \( \mathcal{A}' \)'s intention to communicate \( p \), or equivalently, can arrive at the corresponding belief. Once he has this belief in \( t \), he chooses to prefer \( p \) to \( p' \).

On the other hand, if he is in \( t' \), he can infer that \( s' \) is factual and come to believe that \( \mathcal{A}' \) intends to communicate \( p' \). On this basis, he forms a preference for \( p' \) over \( p \). Note that I have implicitly assumed that the costs involved in transmitting and inferring these two propositions are roughly the same.

For both agents to have access to these derivations we need some assumptions. First, we assume that, given an ambiguous utterance of \( \varphi \) in \( d \), \( \mathcal{A}' \)'s only possible intentions can be to communicate the propositions in the 'range' of \( \varphi \) – in our case, \( p \) or \( p' \). And this must be common knowledge. Both assumptions follow from the second Minimal Content assumption. This much gives the agents access to the content of the two initial situations, and enables the Gricean recognition of intention required for 'nonnatural' communication.

Getting to the preference ordering from the intentions is also not immediate. It requires the further Gricean assumption of rational cooperation between speaker and addressee. Both \( \mathcal{A} \) and \( \mathcal{B} \) need to have common knowledge of being cooperating rational agents to choose preference orderings that are compatible with \( \mathcal{A}' \)'s intentions. Note that while the preference orderings are internally derived, the assumption of cooperation has to be externally given, at least at this level of analysis. In fact, it may be that Grice's Cooperative Principle can be made precise (in such contexts at least) by identifying it with the requirement that both agents order their preferences according to their mutual beliefs about the speaker's intentions. It is worth pointing out here that Gricean cooperation appears to be necessary for all communicative transfers, not just implicatures.

Thus, at this stage, it is common knowledge that if \( \mathcal{B} \) is in \( t \) the intended and preferred interpretation is \( p \), and if he is in \( t' \) it is \( p' \). Unfortunately,
B has as yet no clue about which situation he is in, he knows only that he is either in t or in t'.

Because B cannot tell where he is in the tree it is not clear what his best choice should be as the optimal action is different for t and t'. Note that the last Circumstantial Assumption allows A to think that it is more likely that A is communicating p and that this is common knowledge between them. But this additional information does not enable him to eliminate the uncertainty involved.

A of course knows that it is the situation t that results from her utterance of \( \varphi \). But A also knows that for B t and t' are in the same information set. In fact, it is easy to see that the CICM assumptions imply that the information represented by the tree above becomes common knowledge between them once A utters \( \varphi \).

As it stands, the derived preference ordering above needs to be strengthened into a von Neumann–Morgenstem (N–M) utility function. Each interpretation can then be assigned numerical values, all (positive) linear transformations of the payoff function being considered equivalent.

What justifies this further assumption? To begin with, there is implicit in the situation we’re looking at the information that, at least roughly, \( v(s, \varphi, p) = v(s', \varphi, p') \), where v is the payoff function. As far as B knows, A has no reason to derive greater satisfaction from conveying p rather than \( p' \) or vice-versa. Misinformation should also carry the same degree of dissatisfaction, independently of the particular misinformation conveyed, at least in the situation we’re looking at. This implies that \( v(s, \varphi, p') = v(s', \varphi, p) \). Though these assumptions are appropriate for our particular game (that is, relative to the Circumstantial Assumptions), it should be evident that there will be many discourse situations in which these assumptions will not hold. But there will be different assumptions that will hold in such circumstances and these will suggest appropriate quantitative assessments. Of course, the first two payoffs are greater than the last two. However, these equalities (or inequalities, as the case may be) do not yet allow us to strengthen the ordinal utility functions to an N–M utility function. We can also make the additional plausible assumption that information should be valued quite highly over misinformation. How much more becomes relevant when we consider values for costs of transmitting particular messages. In general, the cost of an utterance and its corresponding interpretation should be much smaller than the satisfaction or dissatisfaction derived from the content of an utterance. This is a reasonable assumption in many ordinary discourse situations but, again, certainly not in all situations. It is this assumption of relative magnitudes of costs and satisfactions and dissatisfactions that gives us the
additional information that justifies our use of N–M utility functions. Both agents can be assumed to have the same payoff function as the information above is common knowledge.

Accordingly, let the payoffs be +10 and −10 for the information and misinformation respectively. To keep things simple, I’m assuming that both agents’ payoffs agree numerically as well.

This makes B’s dilemma clear. He cannot tell which of t and t' is factual and the payoffs are symmetric and conflicting. If he were to choose p then he would get a payoff of 10 if the situation he was in turned out to be t. But if t' were factual he would end up with a minus 10. And the same problem crops up with a choice of p'. If he were to choose one or the other randomly, say by tossing a (fair) coin, he would get an expected payoff of 0, certainly much lower than his maximum possible payoff!

If there was nothing else to the discourse situation it would not be possible for B to disambiguate φ in d. For this to be possible it is necessary for B to take into account the publicly known fact that A is a rational agent and therefore that A has acted optimally in choosing to utter φ. In order to figure out what makes φ A’s optimal choice B needs to consider what A might have said but chose not to.

In general, B will not know all the choices available to A but he shares with A the fact that L is expressive meaning that there is at least one sentence whose minimal content in d is p, and also that there is a sentence that will express p minimally. To keep things simple assume that in the case of p this sentence is the same as μ above. For p' we can assume that there is a sentence μ', say “Every ten minutes a particular man gets mugged in New York”, whose minimal content in d is p'. Here too, I will assume that both A and B consider the same sentence even though the assumption is not strictly necessary.

This enables B to construct the model depicted in Figure 4 of the structure of their interaction once A has said φ.
Note that the two actions involving $\mu$ and $\mu'$ satisfy the Information Assumptions.

We need to account for the payoffs assigned to the two new paths in the tree above. The only difference between $v(s, \varphi, p)$ and $v(s, \mu, p)$ lies in the costs involved and, given the Communication Assumption of least effort, it is reasonable to say that $v(s, \mu, p) < v(s, \varphi, p)$. Also, the additional cost involved in this longer and more specific sentence is small relative to the difference between information and misinformation, as was remarked above. Assume this payoff is 7. The same reasoning goes through for the path with $\mu'$ and we set $v(s', \mu', p') = 7$.

This is $A$’s model of the structure of their interaction. But because it is constructed from their shared knowledge the entire structure is in fact available to both $A$ and $B$ (to $B$ only after $\varphi$ has been uttered). This makes the structure itself common knowledge (also only after $\varphi$ has been uttered). This structure, denoted by $LG(\varphi)$, is a new type of game that I will call a game of partial information. I will also call it a local game because it will turn out to be part of a larger structure called a global game. Thus, $LG(\varphi)$ is a local game of partial information.

We began this discussion with a view to looking at the possible consequences of $A$’s uttering $\varphi$ in $d$. We now have part of the answer to this question. Upon uttering $\varphi$, $LG(\varphi)$ becomes common knowledge between $A$ and $B$.

$LG(\varphi)$ is the choice structure confronting $B$ and is therefore the structure that $A$ needs to consider before she actually chooses $\varphi$. Is there enough information in $LG(\varphi)$ for $B$ to eliminate $t'$ as a possible situation and so be able to choose $p$? This is $B$’s problem. But it is also therefore
Fig. 5. Local game $LG(\mu)$. 

\(s \rightarrow \mu \rightarrow t \rightarrow p \rightarrow +7\)

\(\text{Fig. 5. Local game } LG(\mu).\)

$A$'s problem, because $A$'s optimal choice will depend in part on whether $B$ has enough information to solve this problem.

I will show later that it is possible for $B$ to eliminate $t'$. The argument for this is intricate and it seems best to discuss it after a complete account of the choice structure. For now assume that $B$ is able to eliminate $t'$ and choose $p$ as his preferred interpretation. Since $LG(\varphi)$ is common knowledge between them it seems reasonable to assume (at least for now) that $A$ has access to this reasoning. In fact, it is necessary to assume (or derive from more basic assumptions) that the solution process is common knowledge between them. This enables $B$ to eliminate $t'$ and enables $A$ to anticipate $B$'s choice of $p$. As a result, both receive a payoff of 10. We will say that the value of $LG(\varphi)$, $v[LG(\varphi)]$, is 10.

$A$'s choice structure is now easy to see. For every $\mu$ in her choice set there is a corresponding local game $LG(\mu)$. For example, $LG(\mu)$ is the trivial game in Figure 5. $LG(\mu)$ clearly has a value of 7.

$A$ has to choose the sentence $\mu$ that yields the highest value $v[LG(\mu)]$. This problem is represented in Figure 6.

If $A$ considers only $\varphi$ and $\mu$ here optimal action is to utter $\varphi$. I will call this larger structure (in which the various local games are embedded) the global game (of partial information) and denote this by $GG$. I have implicitly assumed that every local game has a value and this is something I justify later.

But we now have a fairly complete picture of the interaction that makes it possible for $A$ to communicate $p$ to $B$. I will call this structure, which includes both the global and local games, the Strategic Discourse Model or simply the SDM.

I started by considering why $A$ chooses to say $\varphi$. An I have given an

\(v[LG(\varphi)] = +10\)

\(v[LG(\psi)]\)

\(v[LG(\mu)] = +7\)

\(\text{Fig. 6. Global game } GG.\)
account of $A$'s choice problem that shows why $\varphi$ is $A$'s optimal choice. In so doing I have simultaneously explained why $p$ turns out to be the optimal interpretation for $B$ once he has heard $\varphi$. In fact, it is the joint optimality of their choices that makes them individually optimal. It is this dual, "two-sided" optimizing, or what is called strategic rationality, that makes communication possible.

It is important to point out an important asymmetry in the SDM however. $B$ gets to consider only one local game, the one constructed from $A$'s optimal choice. $A$, on the other hand, has to consider all the local games issuing from her choice set and then choose the best one. It is not necessary for $B$ to make conjectures about all the possible things $A$ might have chosen to consider.

It is now time to justify the two provisional statements made earlier. The first concerns the reasoning that $B$ can employ to eliminate $t'$ upon receiving $\varphi$. The second has to do with the existence of a value for every $LG(\mu)$ that $A$ might consider.

7. SOLVING GAMES RATIONAL AGENTS PLAY

Before turning to solving $LG(\varphi)$ it is worth making a general distinction in the context of our game-theoretic analysis. It is important to distinguish between the model $LG(\varphi)$, the different sorts of interactions $LG(\varphi)$ could be a model of, and consequently the different ways in which the model could be analysed. A similar distinction is explicitly made by Aumann [4] and also by Kreps [25]. I have already made it implicitly by separating the model from its analysis.

Solving a game involves finding a pair of "strategies" (one for each player in a two player game) that is in some sense optimal. It appears that there are many different ways to solve the same abstract game, each way being more or less appropriate depending on the particular interpretation we give to the game. A solution concept that may seem appealing in an economic or political content may not be as appealing in a discourse situation; even two different economic contexts or two different discourse situations may provide different grounds for accepting or rejecting a proposed solution.

One persistent problem with many solution concepts is the existence of multiple solutions. This multiplicity is troublesome in most situations because if two or more strategy pairs are optimal then players may not

---

2 This is an important part of the reason why so many solution concepts have been investigated.
know which strategy to play and choosing different or nonmatching optima may result in a suboptimal outcome.

To take Schelling's example [34], if two people have to meet in New York and are not in a position to communicate with each other, any place in the city would do as long as they both choose the same spot. This is a (coordination) game with multiple "solutions". But such solutions are obviously not particularly helpful in prescribing a course of action. Of course, in situations like these, in the absence of other relevant information (perhaps both players are natives and this is common knowledge between them making Grand Central Station a "salient" spot; or it is common knowledge that they are both tourists which might make the Empire State Building "focal") one should not expect unique solutions. In fact, Lewis [28] uses this nonuniqueness as a necessary conditions in the definition of conventions. It is the existence of multiple rational ways of doing something that makes it worthwhile for the players involved to agree on a convention. But it turns out that most solution concepts also allow many unintuitive and implausible solutions to slip past their restrictions, at least under some interpretations of the abstract game under consideration. Exactly which (and how many) solutions are intuitively warranted in a game seems to depend on other features of the particular context being modelled.³

I made the distinction above to keep open the possibility of using different solution concepts for the same local game \(LG(\varphi)\). The solution concept I use here combines one of the more popular solution concepts called Nash equilibrium with the idea of Pareto dominance.

To spell out the concept of a Nash equilibrium we first need to say what a strategy is. A strategy prescribes actions for a player in all possible situations where she has to act. It is essentially a function from the set of all the decision nodes of a player to a set of actions. For example, the function \(\{(s, \varphi), (s', \varphi)\}\) is one of \(A\)'s strategies in the game \(LG(\varphi)\). \(A\) has only two possible choice situations and a strategy specifies her choices in both of them. Obviously, \(A\) has exactly four strategies in this game. It is important to note that a strategy for \(A\) involves a specification of what she would do in \(s'\) even though she knows that \(s'\) isn't factual. This is

³ In many discourse situations nonuniqueness has in fact a different sort of interpretation, making possible the extraction of information not otherwise available. If a local game has multiple solutions and if \(A\) chooses to play it then some sort of ambiguity is left unresolved in the communication and this may convey to the addressee that the ambiguity was intentional for some reason or other (depending on context). We will consider this and other such information later.
necessary because $B$ needs to consider what $A$ might do in $s'$ and so $A$ needs to consider what $B$ might do if he takes into account the possibility that $A$ might be choosing an action in $s'$.

$B$'s strategies involve a slight complication and with it a small refinement of the rough definition of strategy above. $B$ has four choice situations to consider $t, t'$ and $e, e'$. Since $t$ and $t'$ belong to the same information set, $A$ cannot distinguish between the two and so $B$'s choices at $t$ and $t'$ have to be constrained to be the same. That is, the correct domain for the strategy of a player is not the set of all decision nodes but the set of all information sets. In $A$'s case, the domain of a strategy will contain the singleton information sets $\{s\}$ and $\{s'\}$.

$A$ has the following four strategies:

1. $s \mapsto \phi, s' \mapsto \mu' = (\phi, \mu')$
2. $s \mapsto \phi, s' \mapsto \phi = (\phi, \phi)$
3. $s \mapsto \mu, s' \mapsto \phi = (\mu, \phi)$
4. $s \mapsto \mu, s' \mapsto \mu' = (\mu, \mu')$

And $B$ has the following two strategies:

1. $e \mapsto p, t, t' \mapsto p, e' \mapsto p' = (p, p, p') = p$
2. $e \mapsto p, t, t' \mapsto p', e' \mapsto p' = (p, p', p') = p'$

We can simplify the notation if we agree on a convention that specifies the information sets of a player in some predefined order, say, from top to bottom with respect to the tree. We can further simplify the representation of $B$'s strategies by explicitly mentioning only those decisions that represent a "real" choice. Thus, $B$'s choices are constrained to be $p$ and $p'$ at $e$ and $e'$ respectively, so we need not mention these explicitly. Note that we have already made an implicit simplification in our specification
of the strategy functions above. The values of the two functions have been represented as either sentences or propositions. They are actually the corresponding actions, either of uttering the sentence in question or of interpreting the uttered sentence as expressing the relevant proposition.

If $a$ is a strategy of $A$'s and $b$ of $B$'s then the pair $(a, b)$ is called a joint strategy (or a social strategy as against its component individual strategies.) This gives exactly eight strategies in the game $LG(\varphi)$ above. These constitute the strategy space. Intuitively, the unique solution of this game is $(\varphi, \mu', p)$.

What I have defined above is the concept of a pure strategy. In general, players can mix strategies by randomizing on their pure strategy sets. The strategy space then is the set of ordered pairs of probability distributions, one for each player. The argument below can be easily extended to this larger space so I will restrict my remarks to pure strategies. ‘Strategy’ will henceforth mean pure strategy.

A strategy is a Nash equilibrium if no player has an incentive to deviate unilaterally from this strategy. Unilateral deviation by a player is deviation keeping the strategies of other players fixed. Consider $(\mu, \mu', p)$. $B$ certainly has no incentive to deviate from $p$ to $p'$ — it makes no difference which of the two he chooses because $A$’s strategy doesn’t allow the information set $\{t, t'\}$ to become factual. But if $S$ deviates unilaterally to $(\varphi, \mu')$ then she certainly does better. This eliminates $(\mu, \mu', p)$. A quick run through the strategy space will show that only $(\varphi, \mu', p)$ and $(\mu, \varphi, p')$ are Nash equilibria. Call them $N_1$ and $N_2$.

This is probably the most widely used solution concept in the theory of (noncooperative) games. It is worth pointing out though that this criterion cannot be directly deduced from the axioms of utility theory that characterize the behaviour of individual rational agents (see Bernheim [15], Brandenburger and Dekel [16]). Its plausibility lies in its being a necessary condition for rationality if it is already assumed a priori by the players that some rational prescription for action exists in the game. Such an assumption is not always warranted and this has been part of the motivation behind recent efforts to provide more basic foundations for solution concepts (Aumann [5], Kreps [25]).

I will use the Nash criterion without further justification here. We still have to face the fact that there are two Nash equilibria $N_1$ and $N_2$ only

---

Aumann, in particular, has shown that it is possible to deduce the weaker concept of correlated equilibrium from common knowledge of rationality and from the assumption that has come to be called the Harsanyi Doctrine.
one of which is intuitively plausible. We need further conditions to differentiate between $N_1$ and $N_2$.\(^5\)

To solve this multiple equilibrium problem I will use the idea of Pareto-dominance. It says simply that of two strategies in a game, if one results in higher payoffs for all players concerned, the other can be eliminated. Though this appears to make perfect intuitive sense there is a problem with it because it implicitly assumed some degree of correlated action (deviation) among players, something that requires additional assumptions to be warranted in a noncooperative game. In fact, there is often a conflict between the Nash criterion and the Pareto dominance criterion (as evinced, for example, by the Prisoner's Dilemma).

I will use Pareto dominance as a second-order criterion. First, we determine the set of Nash equilibria. Then we apply the Pareto criterion to this set. This ensures that all solutions satisfy the important Nash property that there is no incentive to deviate. That after all is what justifies calling it an "equilibrium" strategy. And this second-order way of using it to eliminate counterintuitive Nash equilibria (and their refinements) is easier to justify. There is also an important general point to be made in justifying this hybrid solution concept and I develop it at some length in a footnote.\(^6\)

\(^5\) Though I will not go into the details here it seems important to point out that none of the standard refinements of Nash equilibrium (sequential, perfect, and proper equilibria, (iterated) dominance, the intuitive criterion, divinity and universal divinity) help here because they all have force by imposing restrictions on "out-of-equilibrium" beliefs. See Kreps [26]. $N_2$ is also a stable equilibrium so the stability requirements of Kohlberg and Mertens [24] do not help here either. An interesting alternative is to see the game as a team-theoretic problem. But solving team problems requires communication among the players. In our case, allowing for preplay communication runs into either or both of two problems. First, there is the danger of an infinite regress because our problem is itself a communication problem. And second, even if such an infinite regress were avoidable, the solution would certainly require a great deal of effort suggesting that languages aren't quite so efficient as they in fact are.

\(^6\) As Schelling [34], most notably, has pointed out, noncooperative games have payoffs that can be thought of (at least loosely) as lying along a continuum ranging from games of pure coordination to games of pure conflict. Intuitively, a game of pure coordination is one in which all players have "perfectly aligned interests", and a game of pure conflict is one in which they have "strictly opposed interests". A sufficient condition for a game to be a pure coordination game is that the payoffs of all players be identical. For a game of pure conflict, it turns out that only two-person zero-sum games can have payoffs with strictly opposed interests. Most games lie in between these two extremes: their payoffs reflect a mix of conflict and coordination, and they are, predictably, called mixed-motive games. The term noncooperative that is used to describe this entire class of games refers not to the "interests" of the players but to the constraints under which they have to choose a rational course of action. Noncooperative games are models of situations in which no "binding agreements" are possible. Cooperative games admit binding agreements between players and it is this element that makes them "cooperative" games, resulting in the possibility of coalition-based
Applying the Pareto criterion to the Nash set ousts $N_2$. The expected payoff from $N_1$ to both players is $0.9(10) + 0.1(7) = 9.7$. The expected payoff to both players from $N_2$ is $0.9(7) + 0.1(10) = 7.3$. This implies that $N_1$ Pareto-dominates $N_2$ and that both players can with certainty choose $N_1$. Since $s$ is factual, this results in $\mathcal{A}$ saying $q$ and in $\mathcal{B}$ choosing $p$ rather

behaviour (Aumann [3]).

The evolution of game theory seems to suggest an implicit bias against an appreciation of the element of coordination that is present in most noncooperative games. All the solution concepts that have been explored appear to focus exclusively on the possibilities for conflictual behaviour. Fully rational behaviour should, at least intuitively, reflect both caution against possible conflictual behaviour by other players as well as the extraction of possibilities for coordination. And the solution concepts we define should capture both sets of possibilities. To be sure, the recent literature shows how subtle considerations of strategic interactions can be. But I'm tempted to think that most game theorists would agree that it is a problem that we do not have a solution concept that gives us the obvious unique solution to the following pure coordination game.

\[
\begin{array}{c|cc}
  & 0,0 & 1,1 \\
 0,0 & & \\
 x, x & & \\
\end{array}
\]

$x = \text{one million}$

Kohlberg and Mertens [24] feel that for the game above with $x = 3$, noncooperative game theory has nothing to offer that will eliminate the equilibrium $(1, 1)$, and that such considerations should fall under the relatively restricted domain of cooperative game theory. However, at least one implication we might draw from the recent interest in criteria that admit correlated behaviour of various sorts (Aumann [5], Kreps and Ramey [27]) is that traditional views of "coordinative" behaviour may not be so obviously true, even if they do turn out to be true in the end. And, if we allow contexts to determine the appropriateness of different solution criteria for the same game, as Kreps has recommended, then there seems little doubt that there is a place for solution concepts that explicitly admit both conflictual and coordinative aspects of the game.

As a first and obvious move towards this I will adopt the mixed "Pareto-Nash" criterion defined above. Note that the game under consideration is a pure coordination game (of partial information) and in such games the set of Pareto-dominant strategies is a subset of the set of Nash equilibria so that there is no conflict between the two solution concepts considered independently. But I will consider more general mixed-motive games below and use the same criterion for these games as well. The Pareto-Nash criterion is admittedly a rather simple concept and is unlikely to have much force. But it should be seen as a first step. The various conflictual solution concepts that have been investigated show much sophistication in exploiting the conflictual elements in games. There does not seem to be any a priori reason to limit the possibilities for sophisticated "cooperative" or coordinative behaviour.

It is worth pointing out explicitly that the entire discussion above is based on three imprecise but suggestive premisses. The first is an intuitive notion of the degree of conflict and coordination in a game. The second is an extremely vague notion of the possibility of separating the conflictual and coordinative aspects of a game. And the third is the identification of the motivations underlying most existing solution concepts with the conflictual aspects of games. It is unclear whether any of these premisses would hold up under more rigorous scrutiny. but we must leave our discussion of these issues here and return to our main story.
COMMUNICATION AND STRATEGIC INFERENCE

than $p'$. (Note that the optimal expected payoff is 9.7, much greater than what is obtainable by tossing a coin (i.e. 0) and foregoing strategic reasoning, at least so long as its costs are ignored.)

This completes our discussion of how the game $LG(\varphi)$ is solved. A strategy that satisfies this solution concept is called a Pareto–Nash equilibrium. $N_1$ turns out to be the unique Pareto–Nash equilibrium $LG(\varphi)$ and so its value $v[LG(\varphi)]$ is 10, as I had asserted above. This is one more place where the difference between games of incomplete information and games of partial information shows up. Not only do they differ in their qualitative structure but also in certain quantitative aspects. The relevant value of the local game for $s_1$ is 10, not 9.7, because $\mathcal{A}$ knows which situation she is in. However, this value is derived from the solution to the local game where the expected value is what counts.

Note that if we assume equal instead of skewed probabilities ("A comet appears every ten years") we are unable to eliminate the second solution $N_2$ and this squares with our intuition as well. In this case the optimal solution would seem to be to spell out the content literally by using $\mu$ as $\varphi$ remains ambiguous. This is interesting because it shows how to justify the use of a more elaborate expression to avoid an ineliminable ambiguity. Also, if we have skewed probabilities as above, but $s'$ is factual rather than $s$, then again the optimal strategy is to spell things out by using $\mu'$ instead of $\mu$.

How do we define the value of a game that has multiple equilibria? In the usual way, as the expected value of the set of multiple values. However, it is unclear what distribution we should use in evaluating expected values. We will assume that, in the absence of any further information each equilibrium strategy is equally likely. (This equiprobable criterion is known to have many weaknesses, but it is perhaps less objectionable in such higher-order contexts.) In the case of the comet, where $s$ and $s'$ have the same likelihood of occurrence, $\mathcal{A}$ should assign equal probabilities to both solutions in the absence of any information about a preference that $\mathcal{B}$ might have for one or the other. Thus, if $\mathcal{B}$ plays $N_1$ they get 10 and if he plays $N_2$ they get $-10$, and the expected value of this set of two values with respect to a uniform distribution is just 0. This is the value that $\mathcal{A}$ should consider in making her optimal decision in the global game.

We need to make certain that every game $LG(\psi)$ that $\mathcal{A}$ might consider does in fact have a value. This is guaranteed to us by a theorem of Nash’s [30] (and its extension by Harsanyi [23] to the existence of Bayesian–Nash equilibria). Every game has at least one Nash equilibrium in the larger space of mixed strategies. In fact, it is easy to show that every game of pure coordination (games in which players have identical payoff functions)
has an equilibrium in pure strategies. In either case, this guarantees in turn that every SDM has a solution. (Actually, the step to the existence of a solution for every SDM isn’t quite so immediate. It requires a consistency condition between the local and global games to be satisfied. This can be found in Parikh [32].)

This completes my analysis of why A chooses to say \( \varphi \) in the utterance situation \( u \), and correspondingly of how B comes to choose the right interpretation \( p \) in the interpretive situation \( i \). But the SDM is not really firmly grounded in the discourse situation as yet. I need to show exactly where this game of partial information is located, whether it is in the “minds” of the agents concerned or whether it is in the ambient circumstances that make up the discourse situation, or both or neither. Situating the SDM will give us a complete model of the communicative transfer between A and B.

We also need to discuss the many assumptions we have made, both the CICM assumptions and a few other assumptions made implicitly along the way. This will suggest some possible refinements of the analysis, and I will develop one of these briefly.

I have also so far been using an intuitive notion of communication. After situating the SDM and discussing the assumptions and some extensions of the model, I will suggest how this notion can be made rigorous in terms of the model.

8. GAMES AND SITUATIONS

I started our analysis of the strategic inference with a discourse situation \( d \) from which we extracted an utterance situation \( u \) and an interpretive situation \( i \). We can extract two more situations from each of \( u \) and \( i \). From \( u \) we extract the choice situation \( c_u \) and the “production” situation \( p_u \). The choice situation is the situation in which A considers her choice structure and \( p_u \) is the situation in which she issues or produces her optimal choice. We have two similar situations for B, a choice situation \( c_i \) in which he considers his choice structure, and an interpretation situation \( i_i \) in which he picks the optimal interpretation. Both \( c_i \) and \( i_i \) are parts of \( i \).

A little situation theory is called for at this point. Situations may be treated as sets of (located) infons. An infon can be modelled as an \((n+2)\)-tuple \( \langle R_n; x_1, \ldots, x_n; i \rangle \) where \( R_n \) is an \( n \)-place relation, \( \{x_1, \ldots, x_n\} \) is its set of arguments, and the \( i \) is a polarity which can be either 1 or 0. If \( i \) is 1, the infon displays the possible fact that the relation \( R_n \) holds of \( \langle x_1, \ldots, x_n \rangle \). If \( i \) is zero, it presents the opposite possibility.
All infons are assumed to be located in some (connected) space-time region \( I \).

A possible situation based on the example is \( a = \{(l; \text{say-}ing; \ A, \ \varphi, \ \mathcal{B}; 1)\} \), where \( l \) might be, say, Riverside Park in New York in 1987. This is a situation in which \( \mathcal{A} \) is saying \( \varphi \) to \( \mathcal{B} \) at \( l \). If the situation describes that slice of the world correctly, we would say that \( a \) is factual and that the single infon in \( a \) is a fact. In general, of course, situations will contain many infons.

One situation is a part of another when it is a subset of the other. For example, \( a \) is a part of \( p_u \). Also, \( d \) has as parts \( u \) and \( i \), which in turn have the parts \( c_u, p_u \) and \( c_i, i_i \), respectively. We will in fact define \( u \) as the union of its two subsets above, as we will \( i \). The discourse situation \( d \) will in our particular example be taken as the union of \( u \) and \( i \). In general, however, it will contain a "sequence" of utterance, (reception) and interpretive situations. More generally yet, it will contain a series of utterances and interpretations intertwined with other actions.

We need to spell out then exactly what the two choice situations contain. \( \mathcal{A} \)'s choice situation \( c_u \) contains \( s \) (from \( LG(\varphi) \)) as a substitution. We had said earlier that \( s \) contains the fact that \( \mathcal{A} \) intends to communicate \( p \). The situation \( c_u \) also contains all the constraints that describe the SDM. For example, it contains the constraint that if \( \mathcal{A} \) says \( \varphi \) \( \mathcal{B} \) will construct \( LG(\varphi) \) in \( c_i \). Another constraint is that if \( \mathcal{A} \) says \( \varphi \) the situation \( t \) will be factual (and a substitution of \( c_i \)). The entire structure of the SDM, the global game and all the local games, are thus embedded in \( c_u \).

\( \mathcal{B} \)'s choice situation is similar. It contains \( t \) as a part though \( \mathcal{B} \) does not know that \( t \) is a part of \( c_i \). It also contains the constraints that capture the local game \( LG(\varphi) \). Note that other situations like \( e \) or \( t' \) never become factual and so never become direct parts of \( c_i \). They enter into the various constraints that make up the SDM and so enter into \( c_u \) and \( c_i \) only indirectly (as "constituents").

Two interesting facts emerge by situating the SDM in this way. The first is that one aspect of the nonwellfoundedness of SDM's (and games in general) becomes explicit. For example, \( c_u \) and \( c_i \) both contain some of the same constraints. Common knowledge of these constraints, which is an additional constraint in both \( c_u \) and \( c_i \) then makes both situations jointly nonwellfounded. (See Barwise [8].)

The second point of interest is that situating the two choice structures (\( \mathcal{A} \)'s and \( \mathcal{B} \)'s) in two separate situations makes it possible to construct in a natural way more general strategic interactions than the ones traditionally modelled in game theory. This more general kind of interaction becomes crucial when we want to define communicative situations and distinguish
them from noncommunicative transfers of information. Essentially, each choice situation contains a model of the interaction between $A$ and $B$. These models may or may not be the same and may or may not be common knowledge. For example, $A$ and $B$ might construct different local games $LG(\phi)$, or $LG(\phi)$ may not be common knowledge. Such situations are not games even though they are interactive choice situations. It is difficult to give any sort of general description of such models because such situations are of many different types. They also raise deep and interesting questions about the nature and meaning of solution concepts. It is important to mention that though they are difficult to model such situations are not uncommon. We will look at a simple instance of such a situation when we consider communicative and noncommunicative transfers of information.

The CICM assumptions also describe situated constraints. As I said above, the Communication and Information Assumptions are located in the background $B$. These background assumptions are common knowledge; we can assume further that they are not part of their so-called "activated" common knowledge, in that agents do not need to consider them explicitly each time they communicate. The Circumstantial Assumptions are constraints in and across $c_u$ and $c_i$.

This completes our situation of the $SDM$. We can see that many of the constraints exist both in the ambient circumstances as well as in the store of knowledge of the agents. This limited embedding of the $SDM$ suffices for our purposes, and has interesting consequences for the concept of communication.

We now have a complete model of the communicative transfer of $p$ from $A$ to $B$, relative to our assumptions. Our next step is to look at some aspects of these assumptions more closely.

9. The Assumptions

Communication is an involved process and thus involves a number of assumptions.

Consider the assumption that agents prefer more relevant information to less (ceteris paribus). It has two features that require comment. First, it is like similar assumptions about various market commodities in economics. It is generally assumed that more of a good is preferred to less. This is reflected in the assumption of concave utility functions (concave in commodity space). However, like money, information is a peculiar second-order type of thing. Though it may be preferred for its own sake, it is typically preferred because it facilitates the satisfaction of other ("first-
order”) goals. This distinction is reflected in the parallel distinction made in statistical decision theory between information-gathering acts or experiments and utility-generating acts or terminal acts (Arrow [2]). It should therefore be derivable form these first-order goals. This fact is reflected in the second feature of note, that it is relevant information that is preferred. I will take the (degree of) relevance of information to be equal to its utility. This provides additional confirmation about the derivability of this assumption from more basic assumptions.

I have also assumed that rational agents minimize effort or cost. The costs associated with a communicative transfer are of various kinds, including costs of processing, memory, and transmission. In the model I implicitly assumed that these costs were shared (more or less) equally. This is unlikely to be true in general, and a more detailed analysis would make a more accurate allocation possible. I will not pursue this however. In the kinds of situations we’re considering, costs are relatively small, so that the SDM is unlikely to be too sensitive to variations in cost allocations.

The third Communication Assumption of cooperation also makes this sharing of costs a reasonable approximation to adopt, especially for the kinds of situations we are looking at. This assumption of cooperation derives essentially from Grice [22]. It is possible to view this crucial assumption as itself a necessary outcome of the demands of rational interaction, however different and potentially conflictual the “ultimate” ends of individuals might be. Arrow [2] cites some reasons for such a possibility in the context of collective action generally. Essentially, collective action (i.e. social “institutions”) can extend the domain of individual rationality in a number of ways. Strawson’s [39] remark that the cooperative assumption is “a precondition of the possibility of the social institution of language” might be seen in this light. Arrow’s nuanced observations of the “invisible” institutions of ethics and morality as an important means for securing the mutual benefits of collective action make relatively weaker claims for the role and status of cooperation. I will leave it at that with three further comments. First, the remarks above touch upon the purely functional or instrumental aspect of these difficult issues. Second, it may be possible to characterize cooperation in a limited range of situations by

---

7 In fact, money may be viewed as a device for summarizing (or perhaps, obviating the need for) a wide range of economy-wide information!

8 One way in which this can happen is through the need for some sufficiently complex social system to mediate the competition for scarce resources; a second is through the possible gains of cooperation, as a result of the nonuniform distribution of capacities and talents in a group, and the nonuniform development of these capacities that occurs as a result of collaborative specialization.
the requirement that both agents order their preferences according to their mutual beliefs about the speaker's intentions. Third, the model identifies relatively precisely why and where cooperation is necessary for communication.

I make two assumptions about the language used by the speaker and addressee. The first of these states that $A$ and $B$ share a language $L$. This is admittedly a little vague, but I have left it so intentionally. It is in fact possible to develop an independent game-theoretic account of what it is to share a language. Lewis [29] gives an outline of one such account. Assuming that the notion of sharing a language can be made precise isn't enough to justify the assumption however. In fact, the assumption that they share an entire language is a strong one and is not strictly necessary for the communication to occur. It also avoids the problem of vagueness. It is obvious that communication is possible with vague languages so that meanings don't need to be fully shared. Also, in general, agents may partially share more than one (vague) language (or communication system). This possibility raises two questions. One is the question of the optimal choice of language for a particular communication, and the other is the deciphering (by the addressee) of what language or code is being used. For our purposes, I will simply assume that $A$ and $B$ fully share just one nonvague language $L$.

The second assumption is that $L$ is a sufficiently expressive language. If it is impossible to express a proposition in some alternative way, then the information communicable via $L$ is limited in important ways. This assumption can be made precise in many ways. First, we might assume that every proposition that can be expressed can be expressed minimally, but this is not strictly necessary. If there is no way to express a proposition minimally, we could do either of two things. One is to consider sentences that express only a proper part of this proposition minimally. This would result in a reduced payoff. In the example, $\mu$ and $\mu'$ led to lower payoffs as a result of higher costs. In general, both higher costs and lower amounts of information may reduce payoffs. Alternatively, we could consider a slightly more complex $SDM$, one in which the minimal content of these alternative expressions would be different from the intended contents, as is the case with $\varphi$. As with the other assumptions I will stick to the simpler versions of this assumption as well.

Finally, all these assumptions need to be common knowledge. And they are all located in $B$. Strictly speaking, some of the considerations above relating to the detailed nature of the payoffs clearly belong in $d$ rather than $B$. But they are true of the full range of discourse situations considered here, so I will relegate them to the background for now.
The Circumstantial Assumptions, as the name suggests, vary with $d$. In the example we considered, the discourse information available to $A$ and $B$ was kept to a minimum. In larger discourse situations, (ones with more $u$'s and $i$'s and other acts), a lot more information is generally available to both agents. In particular, it is possible for $B$ to have some (shared) information about $A$'s initial choice set. This would allow for interesting variations in the game-theoretic structure we've constructed.

A crucial assumption in the analysis is that $A$ and $B$ have available to them shared knowledge of the minimal content of $\varphi$ in $d$. I have also assumed that every utterance has a minimal content, which will in general be different from the intended or communicated content. One can think (loosely) of the minimal content as similar to the literal content. For certain strategic inferences (implicatures) the minimal content is the same as the literal content. Thus, the notion of a minimal content is a generalization of the notion of a literal content. It is possible to develop a simultaneous equation model that incorporates all the strategic inferences involved in understanding an utterance. In such a model, the intended content would be given directly as the solution of the relevant system of equations. Intermediate contents like minimal contents would then be solutions to proper subsets of the full set of equations. What we are doing is abstracting from the larger set of equations and studying just one strategic inference, assuming that we are given the minimal content or solution to the remaining equations that we have abstracted from.

I have also made some assumptions of convenience along the way. One such assumption was that both agents use the same sentences $\mu$ and $\mu'$ in the local games $LG(\varphi)$ they construct. It is easy to see that this assumption is not really necessary. All we need is (common knowledge of the fact) that there is at least one sentence with the relevant properties in the language $L$. This is, of course, guaranteed by the expressiveness of $L$. $A$ and $B$ do not even need to have any particular sentence in mind, as long as they know such a sentence is available (and they know roughly the minimum cost of such sentences). Secondly, the payoff functions that $A$ and $B$ use don't need to be identical, as long as each is a positive linear transformation of the other. Thirdly, the particular numerical probabilities we assigned don't matter, and they can even be different for the two players as long as they fall within a certain range determined by the payoffs.

There were two fundamental assumptions I made implicitly. The first is the assumption that $B$'s interpretive act is publicly observable. Interpretive acts are obviously unobservable; let alone being publicly so. This implies that the payoffs $A$ and $B$ receive at the end of the game are not
publicly observable. This is true of many discourse situations. Often, it will be unclear to the agents involved if the communications was successful or not. In an ongoing discourse, the particular interpretation $B$ chose may be inferrable from subsequent actions by $B$. This suggests some interesting modifications of the model. Just as the contents of utterances are not immediately given, so are the contents of interpretive acts. I will not pursue these possibilities here.

The second basic assumption made implicitly is of the communicative process as an utterance situation followed by an interpretive situation. I consider this in the next section.

10. GAMES AND ATTITUDES

In the $SDM$, $B$ performs the act of interpreting $A$'s utterance. $A$ utters a sentence $\varphi$ and $B$ interprets it. And the structure that enables them to get $p$ across turns out to be a game of partial information.

A complete joint act of communication involves the performance of many separate acts by both speaker and addressee. Austin [7] classified the various speech acts performed by a speaker into locutionary, illocutionary and perlocutionary acts. On the addressee's side, the acts performed as part of a communicative act may be classified (roughly) into acts of perception or reception (of the utterance), interpretive acts (or acts of "understanding" or "securing uptake"), and "responsibe" acts (or acts that are appropriate responses to the message). In the model I implicitly abstracted from the illocutionary act performed by $A$ in saying $\varphi$, namely the act of informing $B$. $B$ cannot be said to have understood $A$'s utterance unless he also understands (at least) this further aspect of $A$'s utterance, called its (illocutionary) force. That is, $B$'s interpretive act should involve not just figuring out the content of $\varphi$ in $d$ but also its force. If we assume that $A$ and $B$ are located in Riverside Park in New York, it is easy to imagine somewhat different Circumstantial Assumptions that make the same utterance a warning.

Typically, the point of performing an illocutionary act in addition to the locutionary act of saying something is to get the addressee to do something, to influence his actions, beliefs or other attitudes. That is, the speaker will want the addressee not simply to interpret her utterance but to perform some further action based on this interpretation. In our example, $A$ will want $B$ to accept the proposition $p$, that is to come to believe $p$. In general, $B$ has the choice of either accepting or rejecting (or discussing) the information that $A$ conveys to him. (More generally, he can accept it probabilistically.) If $A$ had instead said to $B$: "Leave the room", 
A would have, in the appropriate circumstances, performed the (illocutionary) act of requesting B to leave the room, and intended that B leave the room, not simply interpret her utterance. And B would have had the option, for example, of complying with the request or disregarding it.

How does B know what kind of action he is to perform? He needs to identify the (illocutionary) force of the utterance in addition to its content. And once he has done this, he had to decide whether to perform the relevant action intended by A or not.

This requires two modifications of the SDM. Now B performs two actions after hearing (receiving) A's utterance. The first action in interpreting the utterance (i.e. identifying its force and content via a recognition of speaker intentions) and the second is, say, accepting or rejecting it. I will continue to abstract from the first problem of identifying the force, and consider the second action, assuming that the force is given (in much the same way as I assumed that the minimal content was given). This leads to the game in Figure 8.

The notion $F_p$ stands in general for the force with which a proposition has to be interpreted. In the tree above, given our Circumstantial Assumptions, $F_p$ stands for the fact that A is simply conveying the information $p$ to B. The tuple $(p, F_p)$ is the full information communicated.

I have deliberately omitted payoffs. My purpose here is just to point out the kind of game required for a more complete model of communication. Additional assumptions are needed to fill in appropriate payoffs. B may or may not be predisposed to accepting or rejecting A's information (perhaps as a result of his own experiences in NYC). And this predilection of B's may or may not be common knowledge. This situation makes possible a divergence of interests between speaker and addressee. A's goal is presumable to get B to believe a proposition with the least possible effort on her part. But B may need more evidence than just a bald assertion of a fact if he's going to believe it. That is, the least cost sentence may not be the one that is optimal. Earlier, we had a game of pure coordination, one in which both players had identical payoffs. Now we get a game in which the payoff functions for each player may be quite different from each other.

Note that the communicative act itself does not include the particular intended response (accepting a proposition or leaving a room) as a constituent act. The communicative act goes through once B interprets A's utterance correctly. But, in general, it will be necessary for B (and A) to consider the possible actions B might respond with, the corresponding payoff functions, and their shared knowledge of this larger structure in order to communicate successfully.
This kind of (local) game turns out to be similar to what have been called 'signalling games' in the information economics literature (see Kreps [25] for a definition and a bibliography). There are two crucial differences however. One lies in the larger global structure in which these local games are embedded and therefore, also in the interpretation of local games. We call the kinds of games we've constructed games of partial information (instead of incomplete information) to mark this difference. One can see games of incomplete information as special instances of games of partial information, instances in which player $s'$'s (range of possible) choice sets are fully and publicly known before the game starts. A strategic discourse model might then be described as a signalling game of partial information. The other difference is equally, if not more, fundamental. It lies in the fact that playing the game requires agent intentions to be recognized and player payoffs to be endogeneously generated. This is marked by the need...
to make the interpretive act explicit in the game tree, a feature that is conspicuously absent in standard games. This feature carries over to noncommunicative actions as well (in general), and therefore to many situations in economics, not just communicative ones.

I should point out that this analysis does not apply to all communicative acts. As Strawson [40] has emphasized, there is a whole "continuum" of communicative acts ranging from "essentially" conventional to "essentially" nonconventional. Our example is an instance of a nonconventional communicative act. conventional acts (redoubling at bridge, for example) are acts whose intended responses are secured simply by their being performed in the right circumstances, including among other things the fact that the relevant convention be common knowledge between speaker and addressee(s). But what is common to all communicative acts is that both the content and the force of the utterance be understood for successful communication. In some cases this will involve a consideration of addressee response, in others it will not.

It is thus useful to have both types of models, those of pure coordination and those with possible conflict, when we look at various types of communication. Also, the second type of model includes the first type, and when there is no special information about the prior acceptability or otherwise of a proposition to the addressee, the second model is essentially equivalent to the first.

We turn now to an account of communication.

11. Situated Communication

I will assume a familiarity with the Gricean approach to meaning in this section. This includes not only Grice's own accounts [19], [21], [20] but also its various modifications, especially those proposed by Strawson [40], Schiffer [38] and Searle [36]. I will try to indicate how the central insights of this approach can be accommodated in a natural way in my account of situated communication.

The many complexities of the problem (e.g. Grice or Schiffer) as well as the incomplete development of the SDM precludes anything like a complete account of communication. My strategy will be to stick to the example above and reduce, more or less informally, some of the key features of such an account.

Roughly, there are three sorts of constraints that are necessary and sufficient for communication. The first (and relatively obvious, when stated so vaguely) constraint is that communication be a genuinely interactive process. The second constraint is that certain aspects of this interac-
It should be clear from the detailed development of the example how the SDM provides a model of a "genuinely" interactive process. Essentially, the interaction must be strategic, in the precise sense given to this concept by game theory and by the SDM.

I will consider the role of common knowledge in some detail via a well-known counterexample by Strawson [40], or rather, Schiffer's more complete version of it. Suppose A wants B to think that the house he is thinking of buying is rat-infested. A decides to bring about this belief in B by letting loose a rat in the house. She knows that B is watching her and knows that B believes that A is unaware that B is watching her. A intends B to infer, wrongly, from the fact that she let the rat loose that she did so with the intention that B should see the rat, take the rat as "natural" evidence and infer therefrom that the house is rat-infested. A further intends B to realize that the presence of the rat cannot be taken as genuine evidence; but A knows that B will think that A would not so contrive to get B to believe the house is rat-infested unless A had good reasons for thinking it was, and so intends B to infer that the house is rat-infested from the fact that A is letting the rat loose with the intention of getting B to believe that the house is rat-infested.

It should be clear, intuitively, that this is not a case of communication (or even of attempted communication). There is, however, a (noncommunicative) transfer or flow of information from A to B and we can construct a strategic discourse model for this flow in much the same way as we did for the communicative transfer. This is made significantly easier by our situation of A's and B's choice structure in two distinct situations, u and i. This is a discourse situation in which A and B have different models of their interaction.

Consider the two trees, $T_1$ and $T_2$, in Figures 9 and 10 respectively, where $p$ stands for the proposition that the house is rat-infested and $p'$ is its negation.

$T_1$ represents the proposition that a rat appears "naturally" in the house, B interprets this situation as a sign that the house is rat-infested
and has a choice of accepting or rejecting this proposition \( p \). The payoffs are as they are because it would be irrational not to accept such natural evidence as indicating \( p \). \( T_1 \) is not common knowledge (or mutual belief) of course. According to the story, \( B \) believes that \( A \) believes that \( B \) believes the proposition represented by \( T_1 \).

\( T_2 \) is incompletely specified. It represents two sets of choices, one for \( A \) and one for \( B \). \( A \) can choose to "utter" \( \varphi \), which is to let the rat loose. Or \( A \) can choose another action \( \mu \). It is not obvious what other action(s) \( A \) might consider. One possibility is for \( A \) not to do anything, in which case there is nothing for \( B \) to interpret and the payoff \( x \) is 0. Another possibility is for \( A \) to tell \( B \) that the house is rat-infested. Schiffer's story compels us to assume that for some (unspecified) reason or other, this communicative action is relatively inefficient. (As indicated in the previous section, perhaps mere telling may not suffice to get \( B \) to believe \( p \), but a costlier effort might.) That is, better payoffs are possible if \( A \) chooses \( \varphi \) instead of an appropriate linguistic \( \mu \). We need not concern ourselves with the nature of these possibilities as long as we assume that any other \( \mu \) yields a payoff \( x \) that is less than at least one payoff resulting from \( \varphi \).

If \( A \) chooses \( \varphi \), \( B \) can infer either \( p \) or its negation \( p' \), and once again he has the choice of accepting or rejecting either proposition.

The information in the story that determines the payoffs here should be located in the initial situation \( s \). \( A \)'s intention here is that \( B \) infer \( p \) (rather than \( p' \)) by wrongly taking the rat's presence as natural evidence, that is, by assuming \( T_1 \) to be the case rather than \( T_2 \). If this were all, \( B \) should be inclined to reject rather than accept \( p \). But it is further assumed that \( A \) knows that \( B \) knows that \( A \) would not expend such efforts without good reason, hence that knowledge of \( A \)'s intention rather than the actual presence of the rat should be \( B \)'s reason for accepting \( p \). This knowledge of \( B \)'s is what determines the payoffs. Accepting \( p \) has a high (relative) payoff of 10 and rejecting it has a low payoff of \(-10 \). Interpreting the
action as conveying \( p' \) has a low payoff in both cases because the flow is misinformational, that is \( B \) would be wrongly interpreting \( A \) as conveying \( p' \).

\( T_2 \) is also not common knowledge. \( B \) believes the proposition represented by \( T_2 \) to be true. \( B \) knows that the proposition expressed by \( T_1 \) is false, but thinks that \( A \) believes that \( B \) believes it to be true. Further, \( B \) also believes that \( A \) does not know that \( B \) believes \( T_2 \) to be true. These three beliefs, all located in \( t \), provide a model of \( B \)'s choice situation \( c_i \). We will call \( B \)'s choice structure \( SDM_B \). Note that \( SDM_B \) contains two distinct trees, with a more or less complex connection between them.

\( A \) knows all three of \( B \)'s beliefs. This makes the proposition represented by \( T_2 \) false as well, because \( A \)'s true intention is not correctly represented. \( A \)'s true intention is that \( B \) have the three beliefs above and reason as above to infer and accept \( p \). Thus, \( A \)'s choice structure, which we will call \( SDM_A \), contains the same two trees \( T_1 \) and \( T_2 \), and a third \( T_3 \) with the same structure as \( T_2 \), but with the initial situation containing \( A \)'s true intension rather than the false intension above.

The SDM for this flow of information is then simply the pair \( \langle SDM_A, SDM_B \rangle \), \( SDM_A \) being situated in \( u \) and \( SDM_B \) in \( i \). The first choice structure is \( A \)'s model of the strategic interaction, the second is \( B \)'s. Note that there is no common knowledge of any choice structure. This SDM is also a new type of interactive choice situation that has not been studied before. It is not a game, not even of partial information because the common knowledge requirement is not satisfied. We will call more general situations strategic interactions. Strategic interactions will be games only when the common knowledge condition is fulfilled. Every discourse situation can, in principle, be modelled as a strategic interaction by an SDM.

Intuitively, it is clear that the rational choice for \( A \) is to do \( \phi \) and for \( B \) to accept \( p \). Thus, \( p \) flows from \( A \) to \( B \) even though the transfer is not a communicative one. I will not attempt to define a formal solution concept for this SDM. When we give up the assumption of common knowledge the strategic interaction becomes quite complex and the first task would be to define the class of these more general choice structures. They range roughly from one extreme in which each agent has minimal knowledge of the other agent’s model (this is like the single-person choice model assumed in the theory of competitive equilibrium in economics) to the other extreme of full common knowledge (assumed in game theory).

Schiffer in particular has devised a number of similar examples of increasing complexity to argue convincingly that nothing short of common
knowledge of the strategic interaction between $A$ and $B$ is adequate for communication. They key point, as Strawson has pointed out, is that there must not be any latent intentions, as in $T_3$ above. Clark and Marshall [17] makes a similar point in a somewhat different context. Such considerations account for our second constraint on communication.

Our last constraint is that communication be situated. We have already indicated how $SDM$'s are situated. In the Gricean approach we are led inexorably (by the considerations cited above) to postulate intentions of increasing complexity. Even in the original definition proposed by Grice a quite complex intention (made up of three subintentions) is required. This so-called $M$-intention is needed to ensure that the flow of information occurs in the right way and for the right reasons. In the Gricean approach agents and their interactions are not assumed to be situated with the result that all the “work” has to be done by the cognitive states of agents.

As Barwise [9] and Perry [33] have pointed out, taking the situatedness of information flows seriously relieves the agents from considering all the relevant aspects of the embedding circumstances. Just as, to take Perry’s example, an agent does not need to have beliefs about the ambient gravitational force when reaching for a glass of water (or even about the distance between the agent and the glass), so communicating agents do not need to have all the complex intentions apparently required of communication. The structure of the embedding circumstances, that is, the $SDM$ does much of the work. The only intention required of $A$ is the intention to communicate a proposition, or if we include a consideration of addressee response, then to evoke a certain type of response by communicating a certain proposition. We will not argue the case in detail here but it should be easy to see that the $SDM$ does in fact satisfy the conditions fulfilled by all the Gricean intentions.

Just as the noncircumsantal account of beliefs overburdens the belief-forming capacities of agents so the unsituated Gricean approach overburdens the intention-forming capacities of agents. As Perry has pointed out, a circumstantial account does not preclude the possibility of an agent having additional and unnecessary beliefs, and similarly, agents are certainly free to have $M$-intentions should they so prefer. But, situated communication is more efficient than unsituated communication and rational agents will naturally exploit their ambient circumstances optimally. Thus, strictly speaking, situated communication is not an additional necessary condition for communication but is a consequence of the rationality of agents. However, once we take the finiteness of agents seriously, exploiting the situatedness of communication may be the only way (rather
than simply the most efficient way) of effecting certain transfers of information. It is with this in mind that we include it as an independent necessary constraint.

We can now indicate what a definition of communication that incorporates all three constraints would look like. Note that every SDM is a strategic interaction. We will assume further that every discourse situation and background pair \( (d, B) \) "induces" an SDM.

We will say that an agent \( A \) communicates something to another agent \( B \) (by producing \( \varphi \) in discourse situation \( d \) and background \( B \)) iff the SDM induced by \( (d, B) \) is a (situated) game of partial information with a unique solution. What \( A \) communicates to \( B \) will be given by the (Pareto-Nash) equilibrium of the SDM.

The concept of nonnatural meaning can now be defined as attempted communication. That is, \( A \) nonnaturally means something (by producing \( \varphi \)) if \( A \) attempts to communicate something to some \( B \) by producing \( \varphi \) (in some \( d \) and \( B \)). There are other problems with the Gricean account that the situated game-theoretic account must also solve, but my (simple) intention here is merely to indicate what such an account must look like, rather than to give a full-fledged formal definition.

Communication is an example of a joint or collective act. The analysis here suggests that other joint acts might also be analysed in similar ways, as situated games. This has interesting consequences for a theory of coordination.

This concludes my justification of the claim that if the CICM assumptions (as well as the Unspecified Assumptions) are satisfied \( A \) will succeed in communicating \( p \) to \( B \).

12. Conclusion

I have shown how communication and other types of information flows can be modelled by a situation-and-game-theoretic structure called the Strategic Discourse Model. The basic concept in this model is that of a strategic inference. The SDM embodies a number of new ideas. The key ones are the concepts of situated games of partial information, (situated) strategic interactions, and situated communication.

Perhaps the most important consequence of this model is that it suggests necessary and sufficient conditions for communication relative to a class of situations. A transfer of information in this class between two rational agents is a communication if and only if the SDM is a situated game of partial information with a unique solution. We can, if we wish, choose to define communication by this condition. It should be possible then to
derive (some version of) the conditions imposed by Grice [21], Strawson [40] and Schiffer [38] in a precise way, as consequences of theorems. We also need to give a more general definition of the class of SDM's that includes strategic interactions that are not games. I analysed one such example in detail to illustrate the kinds of complexities involved. In its current state, game theory does not in fact have definitions or even examples of these more general interactions in which common knowledge of the structure is not assumed.

The content communicated is given by the Pareto–Nash equilibrium of the game. This provides a precise method for determining the content of an utterance. The current theory of solutions in game theory is inadequate, however. Though there have been many sophisticated and subtle analyses of this problem, we are still short of a satisfactory theory. I used the concept of a Pareto–Nash equilibrium in a relatively cavalier way, without deriving it from first principles, and this is an important problem to address in further developments of the SDM.

The SDM gives us a method for deriving what might be called the fundamental equation of semantics from more basic assumptions. The equation is \( FDM_\mathcal{L}(\varphi, d, B) = \mathcal{C} \), where \( \varphi \) is the sentence uttered, \( d \) the discourse situation, \( B \) the background, and \( \mathcal{C} \) the set of communicated contents. \( FDM_\mathcal{L} \) is the mapping from meanings to contents, relative to a language \( \mathcal{L} \). The model is a first step in this direction. A complete SDM would incorporate all the strategic inferences involved in a complete utterance. Mathematically, this would be a system of simultaneous equations. This simultaneous equation idea offers the most promising avenue for further development of the SDM. It is indeed possible to develop an even more general model of communication and of this equation by treating the language \( \mathcal{L} \) itself as a variable, rather than as a fixed parameter. A partial formulation of such a model is being developed in Parikh [31].

An important consequence of the model is that communication is situated. This makes it possible for the speaker and addressee to communicate with very first-order intentions, unlike the infinitely nested ones postulated in the literature. The ambient game provides the rest of the structure. This parallels Perry's [33] observations about situated beliefs. Situated communication, as an interaction between two agents, is efficient in the same way as the situated action of a single agent.

Another consequence is that the shared language used by the speaker and addressee affects the content communicated in direct and indirect

---

9 This is essentially a general equilibrium model of language and communication in the tradition of Arrow and Debreu.
ways. Directly, of course, the language constrains one to choose from a class of sentences. Indirectly, this choice plays a subtler and equally important role because the content communicated depends not just on what sentence was uttered but also on what sentences might have been uttered but were not. It is this element that, in addition to the more obvious roles played by beliefs and intentions, makes strategic inferences and therefore communication have the range and power they have.

I pointed out several ways in which the SDM could be improved. One relatively straightforward extension is to consider mixed strategies and probabilistic inferences. In this more general case the fundamental equation above would also need to be generalised. \( \mathfrak{c} \) would be transferred together with a probability distribution on it. That is, the appropriate object to consider is a set of propositions with an accompanying probability distribution, not just a proposition. (I am, of course, ignoring the illocutionary force of the utterance here.)

There are several other interesting modifications possible. I will not repeat these here; I simply mention them as interesting directions for further exploration. My principal effort has been to combine ideas from two quite different disciplines and styles of thought in order to develop tools, indeed a language, for studying the various problems connected with the concept of communication, and the flow of information more generally.

The approach developed here can be applied to a wide range of information transfers. Amongst others, it can be used to model problems of reference determination, problems of ambiguity, implicatures, and direct and indirect speech acts. It is interesting that essentially the same considerations apply to all these different types of information flow. If we combine such applications with our definition of communication, we get a new and more precise way to see the intimate connection between Grice's idea of nonnatural meaning and his idea of implicature.

References


1472 Sand Hill Road #305
Palo Alto, CA 94304
U.S.A.